Trust Frameworks & Asymptotic Identity Proofing: A Systems Approach

David Hartzband, D.Sc.
Lecturer, Engineering Systems Division
Massachusetts Institute of Technology

©2013 D. Hartzband
NIST NSTIC Program

• White House initiative proposed in April 2011
  – Focused on providing secure & correct cyber-identities in both public & private sectors to enable trusted online transactions
  – NIST tasked with funding initial NSTIC grants
    • Funded five grants in October 2012
    • D. Hartzband, PI for grant entitled *Identity Ecosystem for Patient Centered Coordination of Care*
Why Trusted Identities?

• Dept. of Commerce estimated $226B US ecommerce retail sales in 2012 comprising about 15B retail transactions
  – This does not count social media usage of 10s of billions of interactions

• Cyber-identity = the establishment & maintenance of one of more electronic identities for the purpose of engaging in ecommerce or other online interaction
  – essential for legal online interaction
  – Trusted identities are established through both agreements & technological support of an identity ecosystem
Why Trusted Identities 2

• Total loss from just from fraud in retail ecommerce, etc. was between $800M & $1.1B in 2011, more now...

• Users have little control over their identity information once it is released to a service provider
  – Could be sold or used for purposes other than the user expects

• Large amount of identity information sharing among service providers can lead to data compromise

• Strong cyber-identities one way of reducing these issues
Identity’s Role in Online Privacy & Security

- Most online security today is based on both role & specific identity
  - Access to social media account: authentication to application directory (ID & password)
  - Access to a specific CDR requires authentication to a local directory (ID & password) & authorization through an IRB directory (ID, role, specific research processes)
  - Access to protected healthcare information: authentication (ID, password); authorization (patient consent); NIST LOA3 identity proofing

- Access to private information is increasing requiring some level of identity proofing
Identity Proofing

• Electronic location & verification of identity attributes for a specific user (screen ID)

  – Collection of identity attributes from:
    • local sources (directories, etc.)
    • Remote sources (public DBs, commercial 3rd parties, etc.)

  – Not all attributes have same weight
  – Multiple sources better than single source

  – NIST LOA 3
    • Requires multi-factor authentication & verification of identity attributes
NSTIC Healthcare Pilot

• Two HIEs
  – San Diego Regional Health Information Exchange
    • Part of the San Diego Beacon eHealth Community
    • Most hospitals in SD County & including Kaiser Permanente Southern CA (San Diego), San Diego VA, SD Council of Clinics, UCSD Medical Center, etc.
  – Gorge Health Community (The Dalles, OR)
    • Small, rural HIE the Columbia River Gorge community in OR & WA
Use Case

Diagram showing the process of a Use Case involving Primary Care Physician (or Staff Member), Cardiologist (or Staff Member), and HIE System for each Pilot Site (Optum Insight - 3D Beacon; Medicity Gorge).

1. Log-in to eReferral system
2. Create eReferral
3. Select Direct Message
4. Send eReferral to Cardiologist
5. Log-in to eReferral System
6. Receive eReferral
7. View & annotate eReferral (treatment recommendations)
8. Autolog-in
9. Generate Decision Support
10. View Decision Support
11. Send Direct Message to PCP
12. Review Treatment Recommendations

©2013 D. Hartzband
Identity Issues

• Population of migrant farm workers get healthcare in both HIEs
  – Pick fruit in OR/WA in summer & get care at Gorge
  – Pick fruit/vegetables in winter in San Diego & Imperial Counties, get healthcare at SD Beacon

• Identity of migrant farm workers extremely hard to verify
Identity Issues 2

• Even identities of providers hard to verify
  – Name on license may be different
  – Name associated with NPI might be different
  – Might have several accounts with different names at various institutions
    • One doc at UCSDMC had 41 different accounts with 5 variations of name

• Patient names also very difficult
  – People’s names changes, but records do not
  – Many ethnicities have names with different forms
  – Clerical errors account for a large proportion of ambiguity
Identity Issues - 3

• I have lived in CA twice for 6 & 4 years respectively
• The CA DMV & Kaiser Permanente had the following variations of my name, none of which were connected:
  – David Hartzband (Name I gave Kaiser)
  – David Jacob Hartzband (Name I gave the DMV as they required a middle name or NMN designation)
  – David J. Hartzband
  – David Jay Hartzband (?)
  – David Yakov Hartzband (?)
  – Davide Yakov Arturovitch Hartzband (birth certificate)
  – David Hartz Band (clerical error that became official)
    • For about a year, Kaiser could not access my medical records because I had several names associated with them & the State of CA thought I was David Hartz Band

• What identities are present in cyberspace representing YOU?
Identity Syndication¹

• Use of multiple sources for location & verification of identity attributes allows for verification with higher assurance

• Syndication uses a search algorithm & known attribute sources to locate identity attributes for a specific user on a per request basis

• Both public & private attribute sources used including commercial & professional sources (AMA, LexisNexis,...)

• Specialized server aggregates attributes & algorithm evaluates/verifies relevance

• Probability model used to assess level of assurance

• High probability identity used in resolution of policy for access of information

¹Syndication work done in participation with Resilient Network Systems, the prime contractor on NIST 12D296
Probability Model

• Cumulative Gamma distribution
  – Projection of a Poisson model as a Bayesian function
  – \( f(x;\alpha,\beta) = \frac{\gamma(\alpha,\beta x)}{(\alpha-1)!} \)
    • \( \alpha,\beta \) are the prior & posterior parameters of the Bayesian distribution
    • Expected mean value = variance
      – Sample size small, so variance small
Cumulative Probability Distribution - Bayesian Gamma: Identity Syndication
## Assurance Levels

<table>
<thead>
<tr>
<th>Number of Identity Attributes</th>
<th>Probability (certainty) of Verification – low assurance attributes</th>
<th>Probability (certainty) of Verification – High assurance attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.26</td>
<td>0.45</td>
</tr>
<tr>
<td>4</td>
<td>0.59</td>
<td>0.90</td>
</tr>
<tr>
<td>6</td>
<td>0.80</td>
<td>0.99</td>
</tr>
<tr>
<td>8</td>
<td>0.91</td>
<td>0.999</td>
</tr>
<tr>
<td>10</td>
<td>0.96</td>
<td>0.9999</td>
</tr>
<tr>
<td>13</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>
Systems Perspective...

- Actual network architecture (VPN layered on TCP/IP) allows decentralization of policy definition & enforcement
- Enables fully distributed network-based services to be used natively
- No central points of attack or centralized data stores for synchronization or corruption issues
- Security & Privacy Policies described in general & then specialized per request allows for “mass customization” of function
Future of Identity

- NSTIC aimed at federated &/or aggregated (syndicated) single ID equivalent for eCommerce & public interaction (healthcare, public registrations (DMV,...) etc.) to improve efficiency/effectiveness of transactions, reduce fraud, etc.
  - May or may not be achievable in finite time – many vested interests with different agenda...
- Latest discussions in NSTIC forums:
  - FIPS rights – current regulations require network entities to inform consumers what data & data sources may be used to authenticate, authorize & identity proof them
    - Consumer can refuse to allow data to be used (fine-grained) & entity can refuse to complete transaction
  - Fine-grained Privacy – consumer can specify what data (at the field level) can be used in eTransactions, network entity must make effort to accomodate
  - “Nym Rights” – allowing consumers full anonymity through pseudonyms & other mechanisms, but still maintaining their ability to be authenticated, etc. – not clear what this means yet
“Futurer” Aspects of Identity

- IMHO - Identity as defined currently is just a façade’
  - Even if we can identify the screen presence to a 4 or 5 9’s level of assurance, it still only tells us a name
  - Actual identity is a context that is built up by your ePresence & your life IRW
  - Syndication is a beginning for developing a sharing that context
  - Eventually it is the context & not the name that is important in your electronic interactions (as it is IRW now)